Discussion paper

No. 2002-92

EMPIRICAL LIKELIHOOD BASED HYPOTHESIS TESTING

By John H.J. Einmahl, Ian W. McKeague

October 2002

ISSN 0924-7815

Empirical likelihood based hypothesis testing

John H.J. Einmahl[*] Ian W. McKeague[†]
Tilburg University Florida State University

September 27, 2002

Abstract

Omnibus tests for various nonparametric hypotheses are developed using the empirical likelihood method. These include tests for symmetry about zero, changes in distribution, independence and exponentiality. The approach is to localize the empirical likelihood using a suitable "time" variable implicit in the null hypothesis and then form an integral of the log-likelihood ratio statistic. The asymptotic null distributions of these statistics are established. In simulation studies, the proposed statistics are found to have greater power than corresponding Cramér–von Mises type statistics.

1 Introduction

We develop an approach to omnibus hypothesis testing based on the empirical likelihood method. This method is known to be desirable and natural for deriving nonparametric and semiparametric confidence regions for mostly finite dimensional parameters, see the recent book Owen (2001) for an excellent account and an extensive bibliography on the topic. Just a few of these papers, however, consider problems of simultaneous inference, and none as far as we know has made a detailed study of omnibus hypothesis testing beyond the case of a simple null hypothesis.

[*]Supported in part by a Senior Fulbright Scholarship. Research partially performed at Eindhoven University of Technology.

[†]Research partially supported by NSF grant 9971784.

[1]*AMS 2000 subject classifications.* Primary: 62G10; secondary: 62G20, 62G30.

[2]*Key words and phrases.* Changepoint, distribution-free, exponentiality, independence, nonparametric likelihood ratio, symmetry, two-sample problem.

[3]Running head: Empirical likelihood tests.

Our approach is based on localizing the empirical likelihood using one or more "time" variables implicit in the given null hypothesis. An omnibus test statistic is then constructed by integrating the log-likelihood ratio over those variables. We consider the proposed procedure to be potentially more efficient than corresponding, often used, Cramér-von Mises type statistics. Four nonparametric problems will be studied in this way: testing for symmetry about zero, testing for a change in distribution (and the two-sample problem), testing for independence and testing for exponentiality. These classical problems have been extensively studied in the literature, but use of the empirical likelihood approach in such contexts appears to be new. Actually, in the book Owen (2001) testing for symmetry and testing for independence are described as "Challenges for empirical likelihood", since the standard method does not work properly here. Our localization approach, however, appears to be a convenient adaptation, which makes empirical likelihood suitable for dealing with these fundamental statistical problems as well.

We first recall the case of a simple null hypothesis. Given i.i.d. observations $X_1, \ldots, X_n$ with distribution function F, consider $H_0 : F = F_0$, where F_0 is a completely specified (continuous) distribution function. Define the localized empirical likelihood ratio

$$R(x) = \frac{\sup\{L(\widetilde{F}) : \widetilde{F}(x) = F_0(x)\}}{\sup\{L(\widetilde{F})\}},$$

where $L(\widetilde{F}) = \prod_{i=1}^n (\widetilde{F}(X_i) - \widetilde{F}(X_i-))$. The empirical distribution function F_n attains the supremum in the denominator, and the supremum in the numerator is attained by putting mass $F_0(x)/\,(nF_n(x))$ on each observation $\leq x$ and mass $(1 - F_0(x))\,/\,(n\,(1 - F_n(x)))$ on each observation $> x$. This easily leads to

$$\log R(x) = nF_n(x) \log \frac{F_0(x)}{F_n(x)} + n\,(1 - F_n(x)) \log \frac{1 - F_0(x)}{1 - F_n(x)}$$

and, provided $0 < F_0(x) < 1$,

$$-2 \log R(x) = \frac{n(F_n(x) - F_0(x))^2}{F_0(x)(1 - F_0(x))} + o_P(1) \xrightarrow{\mathcal{D}} \chi_1^2 \tag{1.1}$$

under H_0. This is a special case of Owen's nonparametric version of the classical Wilks's theorem.

For an omnibus test (consistent against any departure from H_0), however, we need to look at $-2 \log R(x)$ simultaneously over a range of x-values. Taking the integral with respect to F_0, leads to the statistic

$$T_n = -2 \int_{-\infty}^{\infty} \log R(x)\, dF_0(x).$$

If instead of integrating in T_n, we took the supremum over all x, we obtain essentially the statistic of Berk and Jones (1979), who showed that their statistic is more efficient in Bahadur's sense than any weighted Kolmogorov–Smirnov statistic. Li (2000) has introduced an extension of Berk and Jones's approach for a composite null hypothesis that F belongs to a parametric family of distributions. In that case, $R(x) = R_\theta(x)$ for a parameter θ, and Li suggests replacing the unknown θ in Berk and Jones's statistic by its maximum likelihood estimator under the null hypothesis.

Clearly T_n is distribution-free and its small sample null distribution can be approximated easily by simulation. Moreover, from (1.1) and a careful application of empirical process theory, it can be shown (cf. the proof of Theorem 1) that

$$T_n \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(t)}{t(1-t)}\, dt$$

under H_0, where B is a standard Brownian bridge. Under H_0, T_n is asymptotically equivalent to the Anderson–Darling statistic

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F_0(x))^2}{F_0(x)(1-F_0(x))}\, dF_0(x)$$

and the limit distribution may be calculated using a series representation of Anderson and Darling (1952).

We investigate statistics of the form T_n for a variety of nonparametric hypotheses beyond the case of a simple null hypothesis. Testing for symmetry around zero can be handled using $F(-x) = 1 - F(x-)$ and localizing at $x > 0$. To test for exponentiality, we localize using the memoryless property of the exponential distribution. Our method also applies to the two-sample problem, and, more generally, to the nonparametric change point problem; in that case, we localize at (x, t) where t is the proportion of observation time before the changepoint. Testing for independent components in a bivariate distribution function F can be handled using $F(x, y) = F(x, \infty)F(\infty, y)$, with localization at (x, y).

The paper is organized as follows. In Sections 2–5 we examine the four nonparametric testing problems mentioned above and derive likelihood ratio test statistics of the form T_n. Using empirical process techniques, we derive the limiting distribution of T_n in each case. Section 6 contains simulation results comparing the small sample performance of each T_n with a corresponding Cramér–von Mises type statistic, Section 7 is discussion, and proofs are collected in Section 8. Tables of selected critical values for T_n are given in the Appendix.

2 Testing for symmetry

Much has been written on testing symmetry of a distribution around either a known or unknown point of symmetry, some recent contributions being Diks and Tong (1999), Mizushima and Nagao (1998), Ahmad and Li (1997), Modarres and Gastwirth (1996), Nikitin (1996a), and Dykstra, Kochar and Robertson (1995). Early papers include Butler

(1969), Orlov (1972), Rothman and Woodroofe (1972), Srinivasan and Godio (1974), Hill and Rao (1977) and Lockhart and McLaren (1985).

Many of the papers cited above consider the case of a known point of symmetry and use a Cramér–von Mises type test statistic. We also assume that the point of symmetry is known, so without loss of generality it is assumed to be zero. Let $X_1, \ldots, X_n$ be i.i.d. with continuous distribution function F. The null hypothesis of symmetry about zero is

$$H_0 : F(-x) = 1 - F(x-), \quad \text{for all } x > 0.$$

The local likelihood ratio statistic is defined by

$$R(x) = \frac{\sup\{L(\widetilde{F}) : \widetilde{F}(-x) = 1 - \widetilde{F}(x-)\}}{\sup\{L(\widetilde{F})\}}, \quad x > 0.$$

As in the Introduction, the unrestricted likelihood in the denominator is maximized by setting $\widetilde{F} = F_n$, the empirical distribution function. The supremum in the numerator can be found by treating $\widetilde{F}$ as a function of $0 \le p \le 1$, where $\widetilde{F}$ puts mass $p/2$ on the interval $(-\infty, -x]$, mass $p/2$ on $[x, \infty)$, mass $1-p$ on $(-x, x)$, with those masses divided equally among the observations in the respective intervals. That is, the masses on the individual observations in the respective intervals are given by

$$\frac{p/2}{n\hat{p}_1}, \frac{p/2}{n\hat{p}_2}, \frac{1-p}{n(1-\hat{p})},$$

where $\hat{p} = \hat{p}_1 + \hat{p}_2$, $\hat{p}_1 = F_n(-x)$ and $\hat{p}_2 = 1 - F_n(x-)$. The numerator of $R(x)$ is therefore the maximal value of

$$\left(\frac{p/2}{n\hat{p}_1}\right)^{n\hat{p}_1} \left(\frac{p/2}{n\hat{p}_2}\right)^{n\hat{p}_2} \left(\frac{1-p}{n(1-\hat{p})}\right)^{n(1-\hat{p})},$$

which is easily seen to be attained at $p = \hat{p}$. We thus obtain

$$\begin{aligned} \log R(x) &= n\hat{p}_1 \log \frac{\hat{p}}{2\hat{p}_1} + n\hat{p}_2 \log \frac{\hat{p}}{2\hat{p}_2} \\ &= nF_n(-x) \log \frac{F_n(-x) + 1 - F_n(x-)}{2F_n(-x)} \\ &\quad + n(1 - F_n(x-)) \log \frac{F_n(-x) + 1 - F_n(x-)}{2(1 - F_n(x-))}, \end{aligned} \tag{2.1}$$

where $0 \log(a/0) = 0$. Consider as test statistic

$$\begin{aligned} T_n &= -2 \int_0^\infty \log R(x)\, d\{F_n(x) - F_n(-x)\} \\ &= -2 \int_0^\infty \log R(x)\, dG_n(x), \end{aligned}$$

where G_n is the empirical distribution function of the $|X_i|$. Alternatively, we may write

$$T_n = -\frac{2}{n}\sum_{i=1}^{n} \log R(|X_i|).$$

Clearly, T_n is distribution-free; selected critical values are provided in Table A1. The limit distribution of T_n is given by the following result.

Theorem 1 Let F be continuous. Then, under H_0

$$T_n \quad \xrightarrow{\mathcal{D}} \quad \int_0^1 \frac{W^2(t)}{t}\, dt$$

where W is a standard Wiener process.

3 Testing for a changepoint

The nonparametric changepoint testing problem has an extensive literature; recent contributions include Gombay and Jin (1999), Aly (1998), Aly and Kochar (1997), Ferger (1994, 1995, 1996, 1998), McKeague and Sun (1996), and Szyszkowicz (1994). We consider the non-sequential (retrospective) situation with "at most one change", see, e.g., Csörgő and Horváth (1987) and Hawkins (1988).

Let $X_1, \ldots, X_n$ be independent, and assume that for some $\tau \in \{2, \ldots, n\}$ and some continuous distribution functions F, G

$$X_1, \ldots, X_{\tau-1} \sim F \quad \text{and} \quad X_\tau, \ldots, X_n \sim G,$$

with τ, F and G unknown. We wish to test the null hypothesis of no changepoint, $H_0 : F = G$. Define the local likelihood ratio test statistic

$$R(t,x) = \frac{\sup\{L(\widetilde{F}, \widetilde{G}, \tau) : \widetilde{F}(x) = \widetilde{G}(x), \tau = [nt]+1\}}{\sup\{L(\widetilde{F}, \widetilde{G}, \tau) : \tau = [nt]+1\}}$$

for $1/n \le t < 1$ and $x \in \mathbb{R}$, with

$$L(\widetilde{F}, \widetilde{G}, \tau) = \prod_{i=1}^{\tau-1}(\widetilde{F}(X_i) - \widetilde{F}(X_i-)) \prod_{i=\tau}^{n}(\widetilde{G}(X_i) - \widetilde{G}(X_i-)).$$

Set $n_1 = [nt]$, $n_2 = n - [nt]$, and let F_{1n} and F_{2n} be the empirical distribution functions of the first n_1 observations, and last n_2 observations, respectively. Let F_n be the empirical distribution function of the full sample, so $F_n(x) = (n_1 F_{1n}(x) + n_2 F_{2n}(x))/n$. Then

$$\begin{aligned} \log R(t,x) \quad &= \quad n_1 F_{1n}(x) \log \frac{F_n(x)}{F_{1n}(x)} + n_1(1 - F_{1n}(x)) \log \frac{1 - F_n(x)}{1 - F_{1n}(x)} \\ &\quad + n_2 F_{2n}(x) \log \frac{F_n(x)}{F_{2n}(x)} + n_2(1 - F_{2n}(x)) \log \frac{1 - F_n(x)}{1 - F_{2n}(x)}, \qquad (3.1) \end{aligned}$$

where $0\log(a/0) = 0$. Consider as test statistic

$$\begin{aligned} T_n &= -2\int_{1/n}^{1}\int_{-\infty}^{\infty}\log R(t,x)\,dF_n(x)\,dt \\ &= -\frac{2}{n}\sum_{i=1}^{n}\int_{1/n}^{1}\log R(t,X_i)\,dt. \end{aligned}$$

Clearly, T_n is distribution-free; selected critical values are provided in Table A2. The limit distribution of T_n is given by the following result. Let W_0 be a 4-sided tied-down Wiener process on $[0,1]^2$ defined by $W_0(t,y) = W(t,y) - tW(1,y) - yW(t,1) + tyW(1,1)$, where W is a standard bivariate Wiener process.

Theorem 2 Let F and G be continuous. Then, under H_0

$$T_n \xrightarrow{\mathcal{D}} \int_0^1\int_0^1 \frac{W_0^2(t,y)}{t(1-t)y(1-y)}\,dy\,dt.$$

Note that the classical two-sample problem could be handled in a similar way; see Einmahl and Khmaladze (2001) for recent progress on this problem along other lines and for the references therein. We will briefly describe the two-sample problem here, but we will not provide a proof for this case, since it is similar to but easier than the proof for the changepoint problem given in Section 7.

Let $X_1,\ldots,X_n$ be independent, and suppose $X_1,\ldots,X_{n_1}$ $(1 \le n_1 < n)$ have common continuous distribution function F, and $X_{n_1+1},\ldots,X_n$ have common continuous distribution function G; here F and G are unknown. We wish to test the null hypothesis of equal distributions, $H_0 : F = G$. Define the local likelihood ratio test statistic

$$R(x) = \frac{\sup\{L(\widetilde{F},\widetilde{G}) : \widetilde{F}(x) = \widetilde{G}(x)\}}{\sup\{L(\widetilde{F},\widetilde{G})\}},\ x \in \mathbb{R},$$

with

$$L(\widetilde{F},\widetilde{G}) = \prod_{i=1}^{n_1}(\widetilde{F}(X_i) - \widetilde{F}(X_i-))\prod_{i=n_1+1}^{n}(\widetilde{G}(X_i) - \widetilde{G}(X_i-)).$$

Let F_{1n} and F_{2n} be the empirical distribution functions of the first n_1 and last $n_2 := n - n_1$ observations, respectively, and let F_n be the empirical distribution function of the pooled sample $X_1,\ldots,X_n$. Then $\log R(x)$ is equal to the right hand side of (3.1). Consider as a test statistic

$$T_n = -2\int_{-\infty}^{\infty}\log R(x)\,dF_n(x) = -\frac{2}{n}\sum_{i=1}^{n}\log R(X_i)\,;$$

again T_n is distribution-free.

Theorem 2a Let F and G be continuous and assume $n_1, n_2 \to \infty$, as $n \to \infty$. Then, under H_0

$$T_n \xrightarrow{\mathcal{D}} \int_0^1 \frac{B^2(y)}{y(1-y)}\, dy \ ,$$

with B a standard Brownian bridge.

4 Testing for independence

The wide variety of tests for independence has been surveyed by Martynov (1992, Section 12). Here we consider a test for the independence of two random variables.

Let $X_1, \ldots, X_n$ be i.i.d. bivariate random vectors with distribution function F and continuous marginal distribution functions F_1 and F_2. We wish to test the null hypothesis of independence:

$$H_0 : F(x,y) = F_1(x)F_2(y), \text{ for all } x, y \in \mathbb{R}.$$

Define the local likelihood ratio test statistic

$$R(x,y) = \frac{\sup\{L(\widetilde{F}) \colon \widetilde{F}(x,y) = \widetilde{F}_1(x)\widetilde{F}_2(y)\}}{\sup\{L(\widetilde{F})\}}$$

for $(x,y) \in \mathbb{R}^2$, with $L(\widetilde{F}) = \prod_{i=1}^n \widetilde{P}(\{X_i\})$, where $\widetilde{P}$ is the probability measure corresponding to $\widetilde{F}$. Then

$$\begin{aligned}\log R(x,y) &= nP_n(A_{11}) \log \frac{F_{1n}(x)F_{2n}(y)}{P_n(A_{11})} + nP_n(A_{12}) \log \frac{F_{1n}(x)(1-F_{2n}(y))}{P_n(A_{12})} \\ &+ nP_n(A_{21}) \log \frac{(1-F_{1n}(x))F_{2n}(y)}{P_n(A_{21})} + nP_n(A_{22}) \log \frac{(1-F_{1n}(x))(1-F_{2n}(y))}{P_n(A_{22})},\end{aligned}$$

where P_n is the empirical measure, F_{1n} and F_{2n} are the corresponding marginal distribution functions, and

$$\begin{aligned} A_{11} &= (-\infty, x] \times (-\infty, y], \\ A_{12} &= (-\infty, x] \times (y, \infty), \\ A_{21} &= (x, \infty) \times (-\infty, y], \\ A_{22} &= (x, \infty) \times (y, \infty). \end{aligned}$$

Consider as test statistic:

$$T_n = -2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \log R(x,y)\, dF_{1n}(x) dF_{2n}(y).$$

Clearly, T_n is distribution-free; selected critical values are provided in Table A3. The limit distribution of T_n is given by the following result.

Theorem 3 Let F_1, F_2 be continuous. Then, under H_0

$$T_n \xrightarrow{\mathcal{D}} \int_0^1 \int_0^1 \frac{W_0^2(u,v)}{u(1-u)v(1-v)}\, du\, dv$$

where W_0 is a 4-sided tied-down Wiener process on $[0,1]^2$.

The limit distribution above agrees with that in the changepoint problem.

5 Testing for exponentiality

In this section we develop a likelihood ratio based test for exponentiality motivated by the memoryless property of the exponential distribution. Cramér–von Mises type tests based on this property have been proposed by Angus (1982) and Ahmad and Alwasel (1999); we refer to these papers for references to the earlier literature.

Let $X_1, \ldots, X_n$ be i.i.d. non-negative random variables with distribution function F, $F(0-) = 0$, and survival function $S = 1 - F$. Consider the null hypothesis

$$H_0\colon S(x) = \exp(-\lambda x),\ \ x \geq 0,\ \text{for some } \lambda > 0.$$

The local likelihood ratio statistic based on the memoryless property of the exponential distribution is

$$R(x,y) \quad = \quad \frac{\sup\{L(\widetilde{S})\colon \widetilde{S}(x+y) = \widetilde{S}(x)\widetilde{S}(y)\}}{\sup\{L(\widetilde{S})\}}$$

for $x > 0$, $y > 0$, where

$$L(\widetilde{S}) = \prod_{i=1}^{n} (\widetilde{S}(X_i-) - \widetilde{S}(X_i)).$$

Let F_n denote the empirical distribution function. It follows by a straightforward calculation that

$$\log R(x,y) = N_1 \log \frac{n(1-a)}{N_1} + N_2 \log \frac{n(a-b)}{N_2} + N_3 \log \frac{nb(1-a)}{N_3} + N_4 \log \frac{nab}{N_4}$$

where $N_1 = nF_n(x \wedge y)$, $N_2 = n(F_n(x \vee y) - F_n(x \wedge y))$, $N_3 = n(F_n(x+y) - F_n(x \vee y))$, $N_4 = n(1 - F_n(x+y))$, and

$$a = \frac{N_2 + N_3 + 2N_4}{n + N_3 + N_4}, \qquad\qquad b = \left(\frac{N_3 + N_4}{n - N_1}\right) a.$$

Consider as test statistic

$$T_n = -2 \int_0^\infty \int_0^\infty \log R(x,y) \hat{\lambda}^2 e^{-\hat{\lambda}(x+y)}\, dx\, dy,$$

with $\hat{\lambda} = n / \sum_{i=1}^n X_i$. This statistic is distribution-free (under H_0, its distribution does not depend on the parameter λ). Selected critical values for T_n obtained by simulation are displayed in Table A4.

The asymptotic null distribution of T_n is given in the following result. Based on this result, selected critical values for the large sample case are presented in the last row of Table A4. Comparison of Tables A1–4 shows that the convergence of T_n is much slower here than in the previous sections.

Theorem 4 Under H_0,

$$T_n \xrightarrow{\mathcal{D}} 2 \int_0^1 \int_t^1 \frac{st}{(1-s)(1+t)} \left\{ \frac{B(st)}{st} - \frac{B(s)}{s} - \frac{B(t)}{t} \right\}^2 ds\, dt,$$

where B is a standard Brownian bridge.

6 Simulation results

In this section we present simulation results comparing the small sample performance of the proposed likelihood ratio statistic T_n with that of a corresponding Cramér–von Mises type statistic C_n. In each case the powers are based on 10,000 samples, and exact critical values are used (see the Appendix for the T_n critical values).

For the symmetry test, we compared T_n with

$$C_n = n \int_0^\infty \{1 - F_n(x-) - F_n(-x)\}^2 \, dG_n(x),$$

cf. Rothman and Woodroofe (1972). The alternatives are $N(0.3, 1)$ and chi-squared centered about the mean.

Table 1. Power comparison of tests for symmetry. Levels $\alpha = 0.05$ for $n = 50$, and $\alpha = 0.01$ for $n = 100$.

Alternative	$n = 50$		$n = 100$	
	T_n	C_n	T_n	C_n
$N(0.3, 1)$	0.539	0.516	0.629	0.600
centered χ_1^2	0.893	0.732	0.988	0.872
centered χ_2^2	0.505	0.433	0.647	0.495
centered χ_3^2	0.322	0.307	0.332	0.297

For the changepoint test, we compared T_n with

$$C_n = n \int_{1/n}^1 \int_{-\infty}^\infty \{F_{1n}(x) - F_{2n}(x)\}^2 \, dF_n(x)\, dt,$$

cf. Csörgő and Horváth (1988).

Table 2. Power comparison of tests for a changepoint, $n = 50$, $\alpha = 0.05$.

		$\tau = 11$		$\tau = 21$	
F	G	T_n	C_n	T_n	C_n
$N(0,1)$	$N(0,16)$	0.210	0.129	0.735	0.356
unif(0,1)	unif(0.3,1.3)	0.512	0.446	0.837	0.661
exp(1)	exp(2)	0.236	0.229	0.418	0.333
exp(1)	exp(3)	0.506	0.479	0.789	0.683

For the test of independence, we compared T_n with

$$C_n = n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \{F_n(x,y) - F_{1n}(x)F_{2n}(y)\}^2 dF_{1n}(x) dF_{2n}(y),$$

cf. Deheuvels (1981) or Martynov (1992, Section 12). The alternatives are bivariate normal with correlation ρ, and $(U, \beta U + V)$, where U, V are iid uniform on (0,1), for various values of ρ and β.

Table 3. Power comparison of tests for independence at level $\alpha = 0.05$.

Alternative	$n = 20$		$n = 50$	
	T_n	C_n	T_n	C_n
$\rho = 0.4$	0.357	0.341	0.761	0.728
$\rho = 0.5$	0.550	0.520	0.937	0.915
$\beta = 0.5$	0.437	0.389	0.904	0.826
$\beta = 0.6$	0.573	0.523	0.974	0.935

For the test of exponentiality, we compared T_n with

$$C_n = n \int_0^{\infty} \int_0^{\infty} \{S_n(x+y) - S_n(x)S_n(y)\}^2 \hat{\lambda}^2 e^{-\hat{\lambda}(x+y)}\, dx\, dy,$$

cf. Angus (1982). We used levels $\alpha = 0.10$ for $n = 20$, and $\alpha = 0.05$ for $n = 30$. The alternatives were chi-squared, log-normal and Weibull. The log-normal distribution with corresponding normal parameters $\mu = 0$ and σ is denoted LN(σ); the Weibull distribution with scale parameter 1 and shape parameter c is denoted Weibull(c).

Table 4. Power comparison of tests for exponentiality.

Alternative	$n = 20$		$n = 30$	
	T_n	C_n	T_n	C_n
χ^2_4	0.675	0.624	0.717	0.678
LN(0.8)	0.638	0.560	0.696	0.618
LN(1.0)	0.227	0.181	0.201	0.144
Weibull(1.5)	0.619	0.588	0.666	0.638

The proposed statistics show consistent improvement over the corresponding Cramér–von Mises statistics in *all* cases.

7 Discussion

We have developed a rather general localized empirical likelihood approach for testing certain composite nonparametric null hypotheses. We use integral type statistics to establish appropriate limit results. These statistics are somewhat related to Anderson–Darling type statistics, but have the advantage that the implicitly present weight function is automatically determined by the empirical likelihood. Clearly our tests are consistent (against all fixed alternatives). The proofs of our main results (see the next section) require delicate arguments concerning weighted empirical processes to handle "edge" effects in the localized empirical likelihood.

Our approach is tractable in the four cases we have examined because the null hypothesis is expressed in terms of a relatively simple functional equation involving the distribution function(s). Another example in which our approach appears to be useful is in testing bivariate symmetry. More complex null hypotheses, however, might be difficult to handle via our localized empirical likelihood technique. In that sense the goodness-of-fit tests for *parametric* models in Li's (2000) extension of Berk and Jones (1979) are complementary to the present paper (but in contrast with our approach the limit distribution is intractable). However, in the case of testing for exponentiality our test is simpler and more natural. For that case both Li's and our approach can be extended to randomly censored data. Li's approach is not applicable to the other cases we considered.

An interesting direction for future research would be to investigate the Bahadur efficiency of T_n. Nikitin (1996a, 1996b) has studied the Bahadur efficiency of various types of sup-norm statistics in the contexts of testing for symmetry and exponentiality, but it is not clear how to handle statistics of the form T_n.

8 Proofs

We use the following general strategy in each proof. First, we establish the limit distribution for a version of the test statistic in which the range of integration is restricted to a region where the integrand can be approximated uniformly in terms of an empirical process; that region is carefully chosen to avoid a "problematic boundary" where the approximation breaks down. Second, we show that the contribution from the part of the test statistic close to the boundary is asymptotically negligible. The first proof is presented in full detail, but to save space we skip some details in subsequent proofs and concentrate on the key points.

Proof of Theorem 1 The problematic boundary is ∞ in this case. For a given $0 < \varepsilon < 1$, split the range of integration in the test statistic into the bounded interval $[0, x_\varepsilon]$ and its complement, where F has mass $1-\varepsilon$ on $[-x_\varepsilon, x_\varepsilon]$ and mass $\varepsilon/2$ on each side, by symmetry. Decompose the test statistic as $T_n = T_{1n} + T_{2n}$, and note that it suffices to show that as

$n \to \infty$

$$T_{1n} \quad = \quad -2\int_0^{x_\varepsilon} \log R(x)\, dG_n(x) \xrightarrow{\mathcal{D}} \int_\varepsilon^1 \frac{W^2(t)}{t}\, dt, \tag{8.1}$$

and

$$T_{2n} \quad = \quad -2\int_{x_\varepsilon}^{\infty} \log R(x)\, dG_n(x) = O_P(\sqrt{\varepsilon}) \tag{8.2}$$

uniformly in ε, see Billingsley (1968, Theorem 4.2). First consider the leading term T_{1n}. From (2.1), a Taylor expansion of $\log(1+y)$ and the Givenko-Cantelli theorem it follows that, almost surely,

$$\begin{aligned}
&\sup_{0<x\le x_\varepsilon} \left| \log R(x) + \frac{n}{8}\left\{-F_n(-x)+1-F_n(x-)\right\}^2 \left(\frac{1}{F_n(-x)} + \frac{1}{1-F_n(x-)}\right)\right| \\
= &\sup_{0<x\le x_\varepsilon} \left| nF_n(-x)\log\left(1+\frac{1-F_n(x-)-F_n(-x)}{2F_n(-x)}\right)\right. \\
&\qquad + n(1-F_n(x-))\log\left(1+\frac{F_n(-x)-(1-F_n(x-))}{2(1-F_n(x-))}\right) \\
&\qquad \left. + \frac{n}{8}\left\{-F_n(-x)+1-F_n(x-)\right\}^2\left(\frac{1}{F_n(-x)}+\frac{1}{1-F_n(x-)}\right)\right| \\
\le &\sup_{0<x\le x_\varepsilon} \left| \frac{n}{24}(1-F_n(x-)-F_n(-x))^3\left(\frac{1}{(F_n(-x))^2} - \frac{1}{(1-F_n(x-))^2}\right)\right| \\
\le &\frac{1}{24}\sup_{0<x\le x_\varepsilon}\left(\sqrt{n}\left\{1-F_n(x-)-(1-F(x-))+F(-x)-F_n(-x)\right\}\right)^2 \\
&\qquad \cdot \sup_{0<x\le x_\varepsilon}\left\{((1-F_n(x-)-F_n(-x))^2\,\frac{F_n(-x)+1-F_n(x-)}{(F_n(-x))^2(1-F_n(x-))^2}\right\}.
\end{aligned}$$

Now by the weak convergence of the empirical process $\sqrt{n}(F_n - F)$, we immediately obtain that this last bound is $O_P(1)\cdot o_P(1) = o_P(1)$.

Set $U_i = F(X_i)$ and let Γ_n be the empirical distribution function of the U_i. Then by the just obtained uniform approximation of $\log R(x)$, we have

$$\begin{aligned}
T_{1n} \quad = \quad &\int_0^{x_\varepsilon}\Big(\frac{n}{4}\{-F_n(-x)+1-F_n(x-)\}^2 \\
&\qquad \left(\frac{1}{F_n(-x)}+\frac{1}{1-F_n(x-)}\right)+o_P(1)\Big)dG_n(x) \\
= \quad &\int_0^{x_\varepsilon}\frac{n}{4}\left\{-\Gamma_n(F(-x))+1-\Gamma_n(F(x)-)\right\}^2 \\
&\qquad \left(\frac{1}{\Gamma_n(F(-x))}+\frac{1}{1-\Gamma_n(F(x)-)}\right)d\{\Gamma_n(F(x))-\Gamma_n(F(-x))\}+o_P(1)
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\varepsilon/2}^{1/2} \frac{n}{4} \{-\Gamma_n(t) + 1 - \Gamma_n((1-t)-)\}^2 \\
&\qquad \left(\frac{1}{\Gamma_n(t)} + \frac{1}{1-\Gamma_n((1-t)-)}\right) d\{\Gamma_n(t) - \Gamma_n(1-t)\} + o_P(1) \\
&= \frac{1}{4}\int_{\varepsilon/2}^{1/2} \left\{\sqrt{n}(t-\Gamma_n(t)) + \sqrt{n}((1-t) - \Gamma_n((1-t)-)\right\}^2 \\
&\qquad \left(\frac{1}{\Gamma_n(t)} + \frac{1}{1-\Gamma_n((1-t)-)}\right) d\{\Gamma_n(t) - \Gamma_n(1-t)\} + o_P(1), \qquad (8.3)
\end{aligned}
$$

where we used the change of variable $t = F(-x)$. Consider the uniform empirical process

$$\alpha_n(t) = \sqrt{n}(\Gamma_n(t) - t), \;\; t \in [0,1].$$

Since α_n converges in distribution to a Brownian bridge, the Skorohod construction ensures almost sure convergence in supremum-norm of a sequence of uniform empirical processes to a Brownian bridge B. That is, keeping the same notation for these new uniform empirical processes,

$$\sup_{0\le t\le 1} |\alpha_n(t) - B(t)| \to 0 \;\; \text{a.s.}$$

The leading term in (8.3) can then be expressed as

$$\frac{1}{2}\int_{\varepsilon/2}^{1/2} \frac{\{-B(t) - B(1-t)\}^2}{t} d\{\Gamma_n(t) - \Gamma_n(1-t)\} + o(1) \;\; \text{a.s.} \qquad (8.4)$$

By the Helly–Bray theorem the main expression in (8.4) converges a.s. to

$$\int_{\varepsilon/2}^{1/2} \frac{\{-B(t) - B(1-t)\}^2}{t}\, dt \overset{\mathcal{D}}{=} \int_{\varepsilon/2}^{1/2} \frac{W^2(2t)}{t}\, dt = \int_{\varepsilon}^{1} \frac{W^2(t)}{t}\, dt.$$

This settles (8.1).

It remains to show that T_{2n} is asymptotically negligible in the sense of (8.2). Decompose

$$T_{2n} = -2\int_{x_\varepsilon}^{V_n \vee x_\varepsilon} \log R(x)\, dG_n(x) - 2\int_{V_n \vee x_\varepsilon}^{\infty} \log R(x)\, dG_n(x) = T_{3n} + T_{4n},$$

where $V_n = \min(-X_{1:n}, X_{n:n})$ and $X_{i:n}$ denotes the i-th order statistic. Using $|\log(1+y) - y| \le 2y^2$ for $y \ge -1/2$, we find that

$$|\log R(x)| \le \frac{n}{2}(-F_n(-x) + 1 - F_n(x-))^2 \left(\frac{1}{F_n(-x)} + \frac{1}{1-F_n(x-)}\right)$$

for all x. This leads to (cf. (8.3))

$$\begin{aligned} T_{3n} &\le \int_{F(-V_n)\wedge\varepsilon/2}^{\varepsilon/2} \{\alpha_n(t)+\alpha_n((1-t)-)\}^2 \\ &\qquad \left(\frac{1}{\Gamma_n(t)}+\frac{1}{1-\Gamma_n((1-t)-)}\right) d(\Gamma_n(t)-\Gamma_n(1-t)) \\ &= \int_{F(-V_n)\wedge\varepsilon/2}^{\varepsilon/2} \frac{\{\alpha_n(t)+\alpha_n((1-t)-)\}^2}{t^{1/2}}\frac{1}{t^{1/2}} \\ &\qquad \left(\frac{t}{\Gamma_n(t)}+\frac{t}{1-\Gamma_n((1-t)-)}\right) d(\Gamma_n(t)-\Gamma_n(1-t)). \end{aligned} \tag{8.5}$$

The following four sequences are bounded in probability:

$$\sup_{0<t<1}\frac{|\alpha_n(t)|}{t^{1/4}}, \qquad \sup_{0<t<1}\frac{|\alpha_n((1-t)-)|}{t^{1/4}},$$
$$\sup_{U_{1:n}\le t\le 1}\frac{t}{\Gamma_n(t)}, \qquad \sup_{1-U_{n:n}\le t\le 1}\frac{t}{1-\Gamma_n((1-t)-)},$$

in the case of the first two by the Chibisov–O'Reilly theorem, and the last two by Shorack and Wellner (1986, p. 404). Using these bounds inside the integrand of (8.5), and noting that $F(-V_n)\ge \max(U_{1:n}, 1-U_{n:n})$, we obtain

$$T_{3n} = O_P(1)\int_0^{\varepsilon/2}\frac{1}{t^{1/2}}\, d(\Gamma_n(t)-\Gamma_n(1-t)).$$

It follows from integration by parts that

$$\begin{aligned} \int_0^{\varepsilon/2}\frac{1}{t^{1/2}}\, d(\Gamma_n(t)-\Gamma_n(1-t)) &= \int_0^{\varepsilon/2}\frac{1}{t^{1/2}}\, d(\Gamma_n(t)+1-\Gamma_n(1-t)) \\ &= \frac{\Gamma_n(\varepsilon/2)+1-\Gamma_n(1-\varepsilon/2)}{(\varepsilon/2)^{1/2}} \\ &\qquad +\frac{1}{2}\int_0^{\varepsilon/2}\frac{\Gamma_n(t)+1-\Gamma_n(1-t)}{t^{3/2}}\, dt. \end{aligned}$$

Since $\sup_{0<t<1}\Gamma_n(t)/t = O_P(1)$, see, e.g., Shorack and Wellner (1986, p. 404), and similarly $\sup_{0<t<1}(1-\Gamma_n(1-t))/t = O_P(1)$, we obtain now that $T_{3n} = O_P(\sqrt{\varepsilon})$.

Finally consider T_{4n}. Note that $R(x)$ is invariant under a sign-change of the observations X_i. Thus it suffices to evaluate T_{4n} in the case that $F_n(V_n)=1$, which holds either for the original observations or for the sign-changed observations. This gives

$$T_{4n} \le -2\int_{V_n}^{\infty} nF_n(-x)\log\frac{1}{2}\, dG_n(x) = O(n)(1-G_n(V_n))^2 = O_P(1/n),$$

uniformly in ε. The last equality can be seen by noticing that the number of $|X_i|$ greater than V_n is bounded above by a geometric random variable with parameter $1/2$. □

Proof of Theorem 2 Write $U_i = F(X_i)$ and let Γ_{1n}, Γ_{2n} and Γ_n be the corresponding empirical distribution functions. Let $0 < \varepsilon < 1/2$. It suffices to show that as $n \to \infty$

$$\begin{aligned} T_{1n} &= -2\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} \log R(t, Q(y))\, d\Gamma_n(y)\, dt \\ &\xrightarrow{\mathcal{D}} \int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} \frac{W_0^2(t,y)}{t(1-t)y(1-y)}\, dy\, dt \end{aligned} \tag{8.6}$$

and

$$T_{2n} = T_n - T_{1n} = O_P(\sqrt{\varepsilon}) \tag{8.7}$$

uniformly in ε. First consider T_{1n}. By a Taylor expansion it readily follows that uniformly for $\varepsilon \le t, y \le 1-\varepsilon$

$$\begin{aligned} -2\log R(t, Q(y)) &= nt(1-t)(\Gamma_{1n}(y) - \Gamma_{2n}(y))^2 \\ &\quad \left(\frac{1-t}{\Gamma_{1n}(y)(1-\Gamma_{1n}(y))} + \frac{t}{\Gamma_{2n}(y)(1-\Gamma_{2n}(y))}\right)(1+o(1)) + o_P(1). \end{aligned}$$

So instead of T_{1n} we consider

$$\begin{aligned} &\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} nt(1-t)(\Gamma_{1n}(y) - \Gamma_{2n}(y))^2 \\ &\quad \left(\frac{1-t}{\Gamma_{1n}(y)(1-\Gamma_{1n}(y))} + \frac{t}{\Gamma_{2n}(y)(1-\Gamma_{2n}(y))}\right) d\Gamma_n(y)\, dt. \end{aligned}$$

Set $Y_n(t,y) = \sqrt{n}t(1-t)(\Gamma_{1n}(y) - \Gamma_{2n}(y))$. From Csörgő and Horváth (1987), see also McKeague and Sun (1996), it follows that there exists a sequence $\{W_{0,n}\}$ of 4-sided tied-down Wiener processes such that

$$P\left(\sup_{n^{-1/2}<t,y<1-n^{-1/2}} |Y_n(t,y) - W_{0,n}(t,y)| > A\frac{(\log n)^{3/4}}{n^{1/4}}\right) \le Bn^{-\delta}$$

for all $\delta > 0$, where $A = A(\delta)$ and B are constants. Hence it suffices to consider

$$\begin{aligned} &\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} \frac{W_{0,n}^2(t,y)}{t(1-t)}\left(\frac{1-t}{\Gamma_{1n}(y)(1-\Gamma_{1n}(y))} + \frac{t}{\Gamma_{2n}(y)(1-\Gamma_{2n}(y))}\right) d\Gamma_n(y)\, dt \\ &\stackrel{\mathcal{D}}{=} \int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} \frac{W_0^2(t,y)}{t(1-t)y(1-y)}\, d\Gamma_n(y)\, dt + o_P(1), \end{aligned}$$

which implies (8.6) by the Helly–Bray theorem.

It remains to prove (8.7). We will only consider the relevant region of the unit square where in addition both y and t are less than or equal to $\frac{1}{2}$, i.e., we assume $\frac{1}{n} \le t \le \varepsilon$ and

$0 < y \le \frac{1}{2}$, or, $\frac{1}{n} \le t \le \frac{1}{2}$ and $0 < y \le \varepsilon$. Denote this L-shaped region by A_ε. The other regions can be handled in the same way, by symmetry. We prove that

$$\iint_{A_\varepsilon} \log R(t, Q(y))\, d\Gamma_n(y)\, dt \quad = \quad O_P(\sqrt{\varepsilon}). \tag{8.8}$$

We will split, in turn, the region A_ε into several subregions. First we consider the case where $\frac{1}{n} \le t \le \frac{1}{n^{3/5}}$ and $\frac{1}{n^{3/8}} \le y \le \frac{1}{2}$. Note that in this region

$$\begin{aligned}
\left| n_1 \Gamma_{1n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{1n}(y)} \right| &\le n^{2/5} \log n, \\
\left| n_1 (1 - \Gamma_{1n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{1n}(y)} \right| &\le n^{2/5} \log n,
\end{aligned}$$

and with arbitrarily high probability, for large n

$$\begin{aligned}
\left| n_2 \Gamma_{2n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{2n}(y)} \right| &\le |2n_2(\Gamma_n(y) - \Gamma_{2n}(y))| \\
&= \left| \frac{2n_2 n_1}{n} (\Gamma_{1n}(y) - \Gamma_{2n}(y)) \right| \le 2n^{2/5}, \\
\left| n_2 (1 - \Gamma_{2n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{2n}(y)} \right| &\le 2n^{2/5}.
\end{aligned}$$

Hence with high probability, for large n

$$\int_{\frac{1}{n}}^{\frac{1}{n^{3/5}}} \int_{\frac{1}{n^{3/8}}}^{\frac{1}{2}} |\log R(t, Q(y))|\, d\Gamma_n(y)\, dt \le \int_{\frac{1}{n}}^{\frac{1}{n^{3/5}}} 3n^{2/5} \log n\, dt \le \frac{3 \log n}{n^{1/5}} \to 0.$$

Now consider the region $\frac{1}{n^{3/8}} \le t \le \frac{1}{2}$ and $0 < y \le \frac{1}{n^{3/5}}$. In this region we have with high probability, for large n

$$\begin{aligned}
\left| n_1 \Gamma_{1n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{1n}(y)} \right| &\le n_1 \Gamma_{1n}(n^{-3/5}) \log n \le n \Gamma_n(n^{-3/5}) \log n \\
&\le 2n^{2/5} \log n, \\
\left| n_2 \Gamma_{2n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{2n}(y)} \right| &\le n \Gamma_n(n^{-3/5}) \log n \le 2n^{2/5} \log n, \\
\left| n_1 (1 - \Gamma_{1n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{1n}(y)} \right| &\le |2n_1(\Gamma_{1n}(y) - \Gamma_n(y))| \le 2n\Gamma_n(n^{-3/5}) \le 4n^{2/5}, \\
\left| n_2 (1 - \Gamma_{2n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{2n}(y)} \right| &\le |2n_2(\Gamma_{2n}(y) - \Gamma_n(y))| \le 2n\Gamma_n(n^{-3/5}) \le 4n^{2/5}.
\end{aligned}$$

Hence with high probability, for large n

$$\int_{\frac{1}{n^{3/8}}}^{\frac{1}{2}} \int_0^{\frac{1}{n^{3/5}}} |\log R(t, Q(y))|\, d\Gamma_n(y)\, dt \le \int_0^{\frac{1}{n^{3/5}}} 5n^{2/5} \log n\, d\Gamma_n(y) \le \frac{6 \log n}{n^{1/5}} \to 0.$$

Next consider the region $\frac{1}{n} \leq t \leq \frac{1}{n^{3/8}}$ and $0 < y \leq \frac{1}{n^{3/8}}$. In this region

$$\left| n_1 \Gamma_{1n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{1n}(y)} \right| \leq n^{5/8} \log(n^{5/8}), \quad \left| n_1 (1 - \Gamma_{1n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{1n}(y)} \right| \leq n^{5/8} \log(n^{5/8}),$$

and with high probability, for large n

$$\begin{aligned} \left| n_2 \Gamma_{2n}(y) \log \frac{\Gamma_n(y)}{\Gamma_{2n}(y)} \right| &\leq 2n^{5/8} \log n, \\ \left| n_2 (1 - \Gamma_{2n}(y)) \log \frac{1 - \Gamma_n(y)}{1 - \Gamma_{2n}(y)} \right| &\leq \left| \frac{2n_2 n_1}{n} (\Gamma_{1n}(y) - \Gamma_{2n}(y)) \right| \leq 2n^{5/8}. \end{aligned}$$

Hence

$$\int_{\frac{1}{n}}^{\frac{1}{n^{3/8}}} \int_0^{\frac{1}{n^{3/8}}} |\log R(t, Q(y))| \, d\Gamma_n(y) \, dt \leq \frac{4n^{5/8} \log n}{n^{3/4}} \leq \frac{4 \log n}{n^{1/8}} \to 0.$$

In order to handle the remaining part of A_ε we need two lemmas. The first one follows rather easily from Inequality 2 on pp. 415–416 of Shorack and Wellner (1986).

Lemma 1 Let $0 < a_n, b_n \leq 1/2$ with $n a_n b_n \to \infty$ as $n \to \infty$. Then for any $\delta > 0$

$$P \left(\sup_{a_n \leq t \leq 1} \left\{ \left(\sup_{b_n \leq y \leq 1} \frac{\Gamma_{1n}(y)}{y} \right) \vee \left(\sup_{b_n \leq y \leq 1} \frac{y}{\Gamma_{1n}(y)} \right) \right\} > 1 + \delta \right) \to 0.$$

The second lemma follows directly from Komlós, Major and Tusnady (1975), in a similar but easier way than in Csörgő and Horváth (1987).

Lemma 2 Under the same conditions as Lemma 1, there exists a sequence $\{W_{0,n}\}$ of 4-sided tied-down Wiener processes such that

$$\sup_{a_n \leq t \leq 1 - a_n} \sup_{b_n \leq y \leq 1 - b_n} \frac{|Y_n(t, y) - W_{0,n}(t, y)|}{(t(1-t)y(1-y))^{1/4}} \xrightarrow{P} 0.$$

We are now prepared to present the remainder of the proof of Theorem 2. Consider the region $\frac{1}{n^{3/5}} \leq t \leq \varepsilon$ and $\frac{1}{n^{3/8}} \leq y \leq \frac{1}{2}$. We have by a Taylor expansion and Lemma 1 that with high probability, uniformly over this region, for large n

$$\begin{aligned} |\log R(t, Q(y))| &\leq n_1 \frac{(\Gamma_n(y) - \Gamma_{1n}(y))^2}{\Gamma_{1n}(y)(1 - \Gamma_{1n}(y))} + n_2 \frac{(\Gamma_n(y) - \Gamma_{2n}(y))^2}{\Gamma_{2n}(y)(1 - \Gamma_{2n}(y))} \\ &= \frac{n_1 n_2^2}{n^2} \frac{(\Gamma_{1n}(y) - \Gamma_{2n}(y))^2}{\Gamma_{1n}(y)(1 - \Gamma_{1n}(y))} + \frac{n_2 n_1^2}{n^2} \frac{(\Gamma_{1n}(y) - \Gamma_{2n}(y))^2}{\Gamma_{2n}(y)(1 - \Gamma_{2n}(y))}. \end{aligned}$$

We only continue with the first term of this sum; the second one is somewhat easier to deal with. By Lemma 1, with high probability and uniformly over the region, the first term is bounded above by

$$\frac{2Y_n^2(t, y)}{ty} \frac{y}{\Gamma_{1n}(y)} \leq \frac{3Y_n^2(t, y)}{ty} = 3 \left(\frac{Y_n(t, y)}{(ty)^{1/4}} \right)^2 \frac{1}{(ty)^{1/2}}.$$

But by Lemma 2

$$
\begin{aligned}
\int_{\frac{1}{n^{3/5}}}^{\varepsilon}\int_{\frac{1}{n^{3/8}}}^{\frac{1}{2}} \left(\frac{Y_n(t,y)}{(ty)^{1/4}}\right)^2 \frac{1}{(ty)^{1/2}}\,d\Gamma_n(y)\,dt &\stackrel{\mathcal{D}}{=} \int_{\frac{1}{n^{3/5}}}^{\varepsilon}\int_{\frac{1}{n^{3/8}}}^{\frac{1}{2}} \left(\frac{W_0(t,y)}{(ty)^{1/4}}\right)^2 \frac{1}{(ty)^{1/2}}\,d\Gamma_n(y)\,dt \\
&\qquad + o_P(1) \\
&= O_P(1)\int_0^{\varepsilon} \frac{1}{t^{1/2}}\,dt + o_P(1) = O_P(\sqrt{\varepsilon}).
\end{aligned}
$$

Finally it remains to consider the region $\frac{1}{n^{3/8}} \le t \le \frac{1}{2}$ and $\frac{1}{n^{3/5}} \le y \le \varepsilon$. This region, however, can be treated in the same way and yields another term of order $O_P(\sqrt{\varepsilon})$. Hence (8.7) is proved. $\square$

Proof of Theorem 3 The proof is somewhat similar to the changepoint case. Set $X_i = (X_{i1}, X_{i2})$ and denote the empirical distribution function of the $(F_1(X_{i1}), F_2(X_{i2}))$ by G_n, with marginals G_{1n} and G_{2n}. Under H_0, the distribution of $(F_1(X_{i1}), F_2(X_{i2}))$ is uniform on the unit square. Write Q_1, Q_2 for the quantile functions corresponding to F_1, F_2. Let $0 < \varepsilon < 1/2$. It suffices to show that as $n \to \infty$

$$
\begin{aligned}
T_{1n} &= -2\int_{\varepsilon}^{1-\varepsilon}\int_{\varepsilon}^{1-\varepsilon} \log R(Q_1(u), Q_2(v))\,dG_{1n}(u)\,dG_{2n}(v) \\
&\xrightarrow{\mathcal{D}} \int_{\varepsilon}^{1-\varepsilon}\int_{\varepsilon}^{1-\varepsilon} \frac{W_0^2(u,v)}{u(1-u)v(1-v)}\,du\,dv
\end{aligned}
\tag{8.9}
$$

and

$$
T_{2n} = T_n - T_{1n} = O_P(\sqrt{\varepsilon}) \tag{8.10}
$$

uniformly in ε. First consider T_{1n}. By a Taylor expansion it readily follows that uniformly for $\varepsilon \le u, v \le 1-\varepsilon$ (replacing (x,y) by $(Q_1(u), Q_2(v))$ in the definition of the A_{jk})

$$
\begin{aligned}
-2\log R(Q_1(u), Q_2(v)) &= \frac{n(P_n(A_{11})P_n(A_{22}) - P_n(A_{12})P_n(A_{21}))^2}{u(1-u)v(1-v)} + o_P(1) \\
&= \frac{n(P_n(A_{11}) - G_{1n}(u)G_{2n}(v))^2}{u(1-u)v(1-v)} + o_P(1) \\
&= \frac{(\alpha_n(u,v) - v\alpha_{1n}(u) - u\alpha_{2n}(v))^2}{u(1-u)v(1-v)} + o_P(1),
\end{aligned}
\tag{8.11}
$$

with $\alpha_n(u,v) = \sqrt{n}(G_n(u,v) - uv)$, $\alpha_{1n}(u) = \sqrt{n}(G_{1n}(u) - u)$, $\alpha_{2n}(v) = \sqrt{n}(G_{2n}(v) - v)$, $0 < u, v < 1$. So instead of T_{1n} we consider

$$
\int_{\varepsilon}^{1-\varepsilon}\int_{\varepsilon}^{1-\varepsilon} \frac{(\alpha_n(u,v) - v\alpha_{1n}(u) - u\alpha_{2n}(v))^2}{u(1-u)v(1-v)}\,dG_{1n}(u)\,dG_{2n}(v),
$$

which, by standard empirical process theory and a multivariate version of the Helly–Bray theorem, converges in distribution to

$$\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^{1-\varepsilon} \frac{(B(u,v)-vB(u,1)-uB(1,v))^2}{u(1-u)v(1-v)}\,du\,dv,$$

where B is a standard bivariate Brownian bridge: a centered Gaussian process with covariance structure $EB(u,v)B(\tilde{u},\tilde{v}) = (u\wedge\tilde{u})(v\wedge\tilde{v}) - uv\tilde{u}\tilde{v}$, $0<u,\tilde{u},v,\tilde{v}<1$. Observing that

$$\{B(u,v)-vB(u,1)-uB(1,v),\ (u,v)\in(0,1)^2\} \stackrel{\mathcal{D}}{=} \{W_0(u,v),\ (u,v)\in(0,1)^2\},$$

completes the proof of (8.9).

It remains to prove (8.10). We will only consider integration over the region

$$B_\varepsilon = \{(u,v)\in(0,1)^2\colon 0<u\le\varepsilon, 0<v\le 1/2,\ \text{or}, 0<u\le 1/2, 0<v\le\varepsilon\},$$

because of symmetry arguments, cf. the way we handled A_ε in the changepoint case. Because of a further symmetry argument, namely the symmetry in u and v, we will further restrict ourselves to the following three regions which clearly cover $\{(u,v)\in B_\varepsilon\colon u\le v\}$:

$$B_{\varepsilon,1} = \{(u,v)\in(0,1)^2\colon 0<u\le\frac{1}{n^{3/5}}, \frac{1}{n^{3/8}}\le v\le\frac{1}{2}\},$$
$$B_{\varepsilon,2} = \{(u,v)\in(0,1)^2\colon 0<u\le v\le\frac{1}{n^{3/8}}\},$$
$$B_{\varepsilon,3} = \{(u,v)\in(0,1)^2\colon \frac{1}{n^{3/5}}<u\le\varepsilon, \frac{1}{n^{3/8}}\le v\le\frac{1}{2}\}.$$

We almost immediately obtain along the lines of the changepoint case

$$\iint_{B_{\varepsilon,1}\cup B_{\varepsilon,2}} |\log R(Q_1(u),Q_2(v))|\,dG_{1n}(u)\,dG_{2n}(v) = o_P(1); \tag{8.12}$$

where we (again) used that

$$\begin{aligned} |P_n(A_{11}) - G_{1n}(u)G_{2n}(v)| &= |P_n(A_{11}) - G_{1n}(u)(1-G_{2n}(v))| \\ = |P_n(A_{21}) - (1-G_{1n}(u))G_{2n}(v)| &= |P_n(A_{22}) - (1-G_{1n}(u))(1-G_{2n}(v))|. \end{aligned}$$

Moreover, here and in the sequel of the proof we use that, uniform over certain classes of rectangles (the A_{jk}), P_n/P converges to 1 in probability. This follows from, e.g., Chapters 2 and 3 of Einmahl (1987).

For $(u,v)\in B_{\varepsilon,3}$ it rather easily follows that with arbitrarily high probability, uniformly over $B_{\varepsilon,3}$, for large n,

$$\begin{aligned} |\log R(Q_1(u),Q_2(v))| &\le \frac{(\alpha_n(u,v)-v\alpha_{1n}(u)-u\alpha_{2n}(v))^2}{u(1-u)v(1-v)} \\ &\le 12\left\{\frac{\alpha_n^2(u,v)}{uv} + \frac{\alpha_{1n}^2(u)}{u} + \frac{\alpha_{2n}^2(v)}{v}\right\}, \end{aligned}$$

cf. (8.11). This yields that indeed

$$-2\iint_{B_{\varepsilon,3}} \log R(Q_1(u),Q_2(v))\,dG_{1n}(u)\,dG_{2n}(v) = O_P(\sqrt{\varepsilon}),$$

and this, in conjunction with (8.12), yields (8.10). □

Proof of Theorem 4 The quantile function of F is $Q(u) = -\log(1-u)/\lambda$, so we have

$$T_n = -4\int_0^1\int_0^v \log R(Q(u),Q(v))\left(\frac{\hat\lambda}{\lambda}\right)^2((1-u)(1-v))^{\frac{\hat\lambda}{\lambda}-1}\,du\,dv,$$

and it suffices to show that

$$\begin{aligned} T_{1n} &= -4\int_\varepsilon^{1-\varepsilon}\int_\varepsilon^v \log R(Q(u),Q(v))\left(\frac{\hat\lambda}{\lambda}\right)^2((1-u)(1-v))^{\frac{\hat\lambda}{\lambda}-1}\,du\,dv \\ &\xrightarrow{\mathcal{D}} 2\int_\varepsilon^{1-\varepsilon}\int_t^{1-\varepsilon} \frac{st}{(1-s)(1+t)}\left\{\frac{B(st)}{st}-\frac{B(s)}{s}-\frac{B(t)}{t}\right\}^2 ds\,dt \end{aligned} \tag{8.13}$$

and

$$T_{2n} = T_n - T_{1n} = O_P(\sqrt{\varepsilon}) \tag{8.14}$$

uniformly in $0<\varepsilon<1/2$.

First consider (8.13). With $S_n(u) = 1 - F_n(Q(u))$, by a Taylor expansion

$$-2\log R(Q(u),Q(v)) = \frac{n(S_n(u)S_n(v)-S_n(u+v-uv))^2}{u(1-u)(1-v)(2-v)}(1+o_P(1)) \tag{8.15}$$

uniformly for $\varepsilon \le u \le v \le 1-\varepsilon$. Writing

$$\begin{aligned} S_n(u)S_n(v) - S_n(u+v-uv) &= S_n(u)(S_n(v)-(1-v)) + (S_n(u)-(1-u))(1-v) \\ &\quad + ((1-u)(1-v) - S_n(1-(1-u)(1-v))) \end{aligned}$$

and using the weak convergence of the uniform empirical process to a standard Brownian bridge B, we see that the r.h.s. of (8.15) converges weakly on $\varepsilon \le u \le v \le 1-\varepsilon$ to

$$\begin{aligned} &\frac{((1-u)(-B(v)) - B(u)(1-v) + B(1-(1-u)(1-v)))^2}{u(1-u)(1-v)(2-v)} \\ &\stackrel{\mathcal{D}}{=} \frac{\big(-(1-u)B(1-v) - (1-v)B(1-u) + B((1-u)(1-v))\big)^2}{u(1-u)(1-v)(2-v)}. \end{aligned} \tag{8.16}$$

Thus, using the change of variables $s = 1-u$, $t = 1-v$, and noting that $\hat\lambda \xrightarrow{P} \lambda$, we see that (8.13) follows directly from (8.15) and (8.16).

The proof of (8.14) follows along the lines of the previous proofs, in particular the proof of the changepoint case. We only note here that results for weighted empirical processes indexed by intervals, especially Theorem 3.3 in Einmahl (1987), are used to complete the proof. □

Appendix

The following tables provide selected critical values for the four proposed test statistics T_n. The values are based on 100,000 samples in each case.

Table A1. Test for symmetry.

	Percentage points			
n	90%	95%	97.5%	99%
10	2.620	3.392	4.272	5.393
15	2.477	3.325	4.195	5.317
20	2.428	3.271	4.138	5.306
30	2.360	3.154	3.989	5.160
50	2.295	3.081	3.902	5.027
100	2.254	3.041	3.880	5.005
150	2.231	3.002	3.836	4.967

Table A2. Test for a changepoint.

	Percentage points			
n	90%	95%	97.5%	99%
10	1.420	1.667	1.899	2.141
15	1.496	1.756	2.024	2.355
20	1.529	1.804	2.074	2.423
30	1.556	1.832	2.111	2.485

Table A3. Test for independence.

	Percentage points			
n	90%	95%	97.5%	99%
10	1.535	1.792	2.020	2.283
15	1.572	1.841	2.103	2.442
20	1.575	1.852	2.126	2.485
50	1.581	1.861	2.154	2.553

Table A4. Test for exponentiality.

	Percentage points			
n	90%	95%	97.5%	99%
10	0.521	0.734	0.969	1.322
15	0.676	0.906	1.148	1.524
20	0.787	1.004	1.242	1.578
30	0.951	1.155	1.370	1.681
60	1.179	1.390	1.611	1.911
120	1.308	1.522	1.747	2.043
300	1.408	1.631	1.855	2.160
∞	1.467	1.679	1.895	2.192

References

Ahmad, I. A. and Alwasel, I. A. (1999). A goodness-of-fit test for exponentiality based on the memoryless property. *J. Roy. Statist. Soc. Ser. B* **61** 681–689.

Ahmad, I. A. and Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *J. Nonparametr. Statist.* **7** 279–293.

Aly, E.-E. A. A. (1998). Change point tests for randomly censored data. In: *Asymptotic Methods in Probability and Statistics* (Ottawa, ON, 1997), 503–513, North-Holland, Amsterdam.

Aly, E.-E. A. A. and Kochar, S. C. (1997). Change point tests based on U-statistics with applications in reliability. *Metrika* **45** 259–269.

Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain "goodness-of-fit" criteria based on stochastic processes. *Ann. Math. Statist.* **23** 193–212.

Angus, J. E. (1982). Goodness-of-fit tests for exponentiality based on a loss-of-memory type functional equation. *J. Statist. Planning Inf.* **6** 241–251.

Berk, R. H. and Jones, D. H. (1979). Goodness-of-fit test statistics that dominate the Kolmogorov statistics. *Z. Wahrsch. Verw. Geb.* **47** 47–59.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Csörgő, M. and Horváth, L. (1987). Nonparametric tests for the changepoint problem. *J. Statist. Plan. Infer.* **17** 1–9.

Csörgő, M. and Horváth, L. (1988). Nonparametric methods for changepoint problems. In: *Handbook of Statistics* **7** (P. R. Krishnaiah and C. R. Rao, eds.), 403–425. Elsevier, Amsterdam.

Deheuvels, P. (1981). An asymptotic decomposition for multivariate distribution-free tests of independence. *J. Multivariate Anal.* **11** 102–113.

Diks, C. and Tong, H. (1999). A test for symmetries of multivariate probability distributions. *Biometrika* **86** 605–614.

Dykstra, R., Kochar, S. and Robertson, T. (1995). Likelihood ratio tests for symmetry against one-sided alternatives. *Ann. Inst. Statist. Math.* **47** 719–730.

Einmahl, J. H. J. (1987). *Multivariate Empirical Processes. CWI Tract 32.* Centre for Mathematics and Computer Science, Amsterdam.

Einmahl, J. H. J. and Khmaladze, E. V. (2001). The two-sample problem in $\mathbb{R}^m$ and measure-valued martingales. *State of the Art in Probability and Statistics, IMS Lecture Notes - Monograph Series* **36** 434-463.

Ferger, D. (1994). Nonparametric change-point tests of the Kolmogorov-Smirnov type. In: *Change-point Problems*, 145–148. IMS Lecture Notes Monogr. Ser., 23, Inst. Math. Statist., Hayward, CA.

Ferger, D. (1995). Nonparametric tests for nonstandard change-point problems. *Ann. Statist.* **23** 1848–1861.

Ferger, D. (1996). On the asymptotic behavior of change-point estimators in case of no change with applications to testing. *Statist. Decisions* **14** 137–143.

Ferger, D. (1998). Testing for the existence of a change-point in a specified time interval. *Advances in Stochastic Models for Reliability, Quality and Safety* (Schierke, 1997), 277–289, Stat. Ind. Technol., Birkhauser Boston, Boston.

Gombay, E. and Jin, X. (1999). Sign tests for change under alternatives. *J. Nonparametr. Statist.* **10** 389–404.

Hawkins, D. L. (1988). Retrospective and sequential tests for a change in distribution based on Kolmogorov-Smirnov-type statistics. *Sequential Analysis* **7** 23–51.

Hill, D. L. and Rao, P. V. (1977). Tests of symmetry based on Cramér–von Mises statistics. *Biometrika* **64** 489–494.

Komlós, J., Major, P. and Tusnady, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Geb.* **32** 111–131.

Li, G. (2000). A nonparametric likelihood ratio goodness-of-fit test for survival data. Preprint.

Lockhart, R. A. and McLaren, C. G. (1985). Asymptotic points for a test of symmetry about a specified median. *Biometrika* **72** 208–210.

Martynov, G. V. (1992). Statistical tests based on empirical processes, and related problems. *J. Soviet Math.* **61** 2195–2271.

McKeague, I. W. and Sun, Y. (1996). Transformations of Gaussian random fields to Brownian sheet and nonparametric change-point tests. *Statist. Probab. Lett.* **28** 311–319.

Mizushima, T. and Nagao, H. (1998). A test for symmetry based on density estimates. *J. Japan Statist. Soc.* **28** 205–225.

Modarres, R. and Gastwirth, J. L. (1996). A modified runs test for symmetry. *Statist. Probab. Lett.* **31** 107–112.

Nikitin, Y. Y. (1996a). On Baringhaus–Henze test for symmetry: Bahadur efficiency and local optimality for shift alternatives. *Math. Methods Statist.* **5** 214–226.

Nikitin, Y. Y. (1996b). Bahadur efficiency of a test of exponentiality based on a loss-of-memory type functional equation. *J. Nonparametr. Statist.* **6** 13–26.

Orlov, A. I. (1972). Testing the symmetry of a distribution. *Theory Prob. Appl.* **17** 372–377.

Owen, A. B. (2001). *Empirical Likelihood.* Chapman & Hall/CRC, Boca Raton.

Rothman, E. N. D. and Woodroofe, M. A. (1972). A Cramér–von Mises type statistic for testing symmetry. *Ann. Math. Statist.* **43** 2035–2038.

Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Srinivasan, R. and Godio, L. B. (1974). A Cramér-von Mises type statistic for testing symmetry. *Biometrika* **61** 196–198.

Szyszkowicz, B. (1994). Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stochastic Process. Appl.* **50** 281–313.

Department of Econometrics & OR
Tilburg University
P.O. Box 90153
5000 LE Tilburg
The Netherlands
E-mail: j.h.j.einmahl@uvt.nl

Department of Statistics
Florida State University
Tallahassee, FL 32306-4330
USA
E-mail: mckeague@stat.fsu.edu